Developmental Science

WILEY

Developmental Science

RESEARCH ARTICLE

A Longitudinal Study of Adolescent-to-Young Adult Executive Function Development in Seven Countries

Ann Folker[1] | Christina Bertrand[1] | Yelim Hong[2] | Laurence Steinberg[3] | Natasha Duell[4] | Lei Chang[5] | Laura Di Giunta[6] | Kenneth A. Dodge[7] | Sevtap Gurdal[8] | Daranee Junla[9] | Jennifer E. Lansford[7] | Paul Oburu[10] | Concetta Pastorelli[6] | Ann T. Skinner[7] | Emma Sorbring[8] | Marc H. Bornstein[11,12,13] | Liliana Maria Uribe Tirado[14] | Saengduean Yotanyamaneewong[9] | Liane Peña Alampay[15] | Suha M. Al-Hassan[16] | Dario Bacchini[17] | Kirby Deater-Deckard[1,18]

[1]University of Massachusetts Amherst, Amherst, Massachusetts, USA | [2]University of Texas Austin, Austin, Texas, USA | [3]Temple University, Philadelphia, Pennsylvania, USA | [4]California Polytechnic State University San Luis Obispo, San Luis Obispo, California, USA | [5]University of Macau, Taipa, Macau, China | [6]Università di Roma “La Sapienza”, Rome, Italy | [7]Duke University, Durham, North Carolina, USA | [8]University West, Trollhättan, Sweden | [9]Chiang Mai University, Chiang Mai, Thailand | [10]Maseno University, Maseno, Kenya | [11]Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland, USA | [12]UNICEF, New York, USA | [13]Institute for Fiscal Studies, London, UK | [14]Universidad de San Buenaventura, Medellín, Colombia | [15]Ateneo de Manila University, Quezon City, Philippines | [16]Abu Dhabi Early Childhood Authority, Abu Dhabi, UAE | [17]University of Naples “Federico II”, Naples, Italy | [18]Helsinki Collegium for Advanced Studies, Helsinki, Finland

Correspondence: Kirby Deater-Deckard (kdeaterdeck@umass.edu)

Received: 27 August 2024 | **Revised:** 4 February 2025 | **Accepted:** 23 May 2025

Funding: This research was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant RO1-HD054805 and Fogarty International Center Grant RO3-TW008141. This research was also supported by National Institute on Drug Abuse Grant P30 DA023026, the Intramural Research Program of the National Institutes of Health/National Institute of Child Health and Human Development, United States, United Nations Children’s Fund, and an International Research Fellowship at the Institute for Fiscal Studies, London, United Kingdom, funded by the European Research Council under the Horizon 2020 research and innovation program (Grant 695300-HKADeC-ERC-2015-AdG).

Keywords: adolescence | executive function | longitudinal | meta-analysis | multi-site | young adult

ABSTRACT

Executive functioning (EF) is an important developing self-regulatory process that has implications for academic, social, and emotional outcomes. Most work in EF has focused on childhood, and less has examined the development of EF throughout adolescence and into emerging adulthood. The present study assessed longitudinal trajectories of EF from ages 10 to 21 in a diverse, international sample. 1093 adolescents (50.3% female) from eight locations in seven countries completed computerized EF tasks (Stroop, Tower of London [ToL], Working Memory [WM]) at ages 10, 14, 17, and 21. Latent growth curve models were estimated to understand the average performance at age 10 and the change in performance over time for each task. Meta-analytic techniques were used to assess the heterogeneity in estimates between study sites. On average, EF task performance improved across adolescence into young adulthood with substantial between-site heterogeneity. Additionally, significant individual differences in EF task performance at age 10 and change in EF task performance over time characterized the full sample. EF improves throughout adolescence into young adulthood, making it a potentially important time for intervention to improve self-regulation.

Ann Folker and Christina Bertrand are co-lead authors.

Summary

- Performance on the Tower of London, Stroop, and Working Memory improves on average in an international sample of 10–17-year-olds.
- For all three executive function tasks, there were significant individual differences in the intercept and slope.
- There was substantial between-site heterogeneity in the slope and intercept; international study sites differ in their starting points and change in executive function task performance.

1 | Introduction

Executive functioning (EF) is a key aspect of self-regulation that also plays an important role in an individual's purposeful goal pursuits. EF is linked to many developmental outcomes spanning behavioral, mental, and physical health as well as academic and interpersonal functioning (Cortés Pascual et al. 2019; Diamond 2016; Hofmann et al. 2012; Jacobson et al. 2011; Nyongesa et al. 2019). However, nearly all existing research regarding EF development has relied on cross-sectional designs, single-location studies, and samples from Western nations. To our knowledge, no study has explored the *longitudinal development* of EF across adolescence in a variety of locations around the world. Cross-national research is important to understand whether EF development is culturally specific or more culturally general. In the current study, we examined age-typical and individual differences in longitudinal growth in EF skills from 10 to 21 years in cities or towns in seven countries that vary on sociodemographic dimensions, including ethnicity, predominant religion, and economic indicators.

EF is defined in a variety of ways. In the current research, we utilized the commonly applied "university and diversity" operationalization of EF (Friedman and Miyake 2017), which includes a general EF factor that is composed of three subcomponents including inhibitory control (i.e., suppressing dominant, often prepotent, responses), updating/working memory (WM) (i.e., maintaining and revising relevant information), and set shifting/attentional control (i.e., moving attention between multiple tasks or dimensions of information). Different experimental tasks are employed as measures for these three fundamental EF components. Typically, inhibition is assessed using the Stroop task, which requires adolescents to inhibit or override a prepotent response (Hofmann et al. 2012). The backward digit span measures adolescents' WM by evaluating their accuracy in repeating a sequence of numbers backward (Wechsler 1945). The Tower of London (ToL) measures attentional control, set shifting, and planning to evaluate an adolescent's ability to inhibit actions before a play is completely developed (Shallice 1982; Steinberg 2017). Studying the development of EF yields insights into how these critical cognitive abilities emerge and evolve over time.

EF develops very rapidly across early childhood. As youth develop across middle childhood and adolescence, EF capacities continue to develop but do so more gradually, and individual differences become stable (Anderson et al. 2010; Ferguson et al. 2021; Icenogle et al. 2019; Raffaelli et al. 2005; Steinberg 2005). However, the patterns of developmental improvement (e.g., shape and rate of change) tend to be modest, and they may differ somewhat depending on how EF is measured and on characteristics of the study samples (e.g., Klimkeit et al. 2004; Lee et al. 2013; Magar et al. 2010; Theodoraki et al. 2020).

Very little longitudinal research examines the development of self-regulation beyond middle childhood. One study that utilized cross-sectional data from adolescents aged 11–17 years found age-related improvements in EF tasks that assessed cognitive regulatory processes, specifically in the areas of updating WM and mental set-switching. This result may indicate linear improvements in certain aspects of self-regulation over the course of adolescence from ages 11 to 17 (Magar et al. 2010). However, response inhibition, as assessed with the go-no-go task, did not show significant change across this age range, a finding that may suggest that adult-level competence in inhibitory control might have been reached before the age of 12 years (Klimkeit et al. 2004; Lee et al. 2013). However, another study, which involved a cross-sectional sample of 14–18-year-olds, reported no evidence of age-related changes in performance on tasks measuring WM or shifting. Nonetheless, significant changes in performance were observed on the inhibition task (Theodoraki et al. 2020). A multinational cross-sectional study involving 10–30-year-olds across 11 countries found consistent increases in self-regulation (measured by Stroop task, self-reported planning, and ToL task) from preadolescence to young adulthood, stabilizing between ages 23 and 26. Although the magnitude of age-related trends varied slightly, the overall developmental patterns were largely consistent across countries (Steinberg et al. 2018). Additionally, a study combining cross-sectional and longitudinal data from four independent datasets (ages 8–35) demonstrated that EF follows a non-linear trajectory, with rapid development from late childhood to mid-adolescence (ages 10–15) and stabilization in late adolescence (ages 18–20) (Tervo-Clemmens et al. 2023). However, it is important to note that most of these studies were based on cross-sectional data, emphasizing the need for further research using longitudinal data to control for potential cohort effects and to be able to estimate individual trajectories of development over time.

In sum, nearly all extant research on the development of EF in adolescence is cross-sectional and based on samples of adolescents from single locations. There is a need for longitudinal studies of adolescents from multiple geopolitical locations and cultures, including communities in the global south that are vastly underrepresented in psychological research (Barrett 2022; Nielsen et al. 2017). Thus, the goal of the present study was to examine individual differences in longitudinal trajectories of EF in adolescents-to-young adults from 10 to 21 years in locations in seven countries spanning five continents. We hypothesized that the indicators of EF would gradually improve across this time period and that this pattern of improvement would be consistent across the international locations of the study.

2 | Method

2.1 | Participants

Participants included 1093 adolescents (50.3% female) from eight locations in seven countries (Medellín, Colombia, $N = 108$; Rome and Naples, Italy, $N = 213$; Zarqa, Jordan, $N = 114$; Kisumu, Kenya, $N = 100$; Quezon City, Philippines, $N = 120$; Chiang Mai, Thailand, $N = 120$; Durham, North Carolina, United States, $N = 318$) in the longitudinal Parenting Across Cultures (PAC) study. Data from annual waves 3, 6, 9, and 12 were used in the present analysis because these waves included EF data. Adolescents were $M = 10.72$ years (SD = 0.67) at wave 3, $M = 14.51$ years (SD = 0.96) at wave 6, $M = 17.75$ years (SD = 1.02) at wave 9, and $M = 21.13$ years (SD = 0.84) at wave 12. See Supplementary Table S1 for detailed description of adolescent sex and age by study site.

Families were recruited to participate through letters sent to schools at each site. For more details about the recruitment process, see Lansford et al. (2014). Economic diversity in the sample was ensured by sampling from private and public schools and including high- to low-income families in representative proportions for each site. However, the samples are not nationally representative for each site.

2.2 | Procedures

The current study utilized task-based measurement of EF, but to be thorough we provide here a general summary of the procedures used in the longitudinal PAC study. Children and their parents completed various measures at each time point via face-to-face interviews (in the adolescent's home, at the child's school, at the research site, or in another location chosen by the families), telephone interviews, via web-based survey, by written questionnaires, and using a laptop computer. All questionnaires/interview questions were forward- and back-translated by translators fluent in English and the target language to clarify any item-by-item ambiguities in linguistic or semantic content (Erkut 2010). During the translation process, translators also noted any items that did not translate well (i.e., were inappropriate, not culturally sensitive, had multiple meanings), and suggested improvements; Maxwell 1996; Peña 2007). The procedures were approved by Institutional Review Boards at each study site. Parents and children provided consent/assent.

For the first, second, and third data collection periods, adolescents completed computerized EF tasks in person or via Zoom with a trained interviewer on a laptop computer using EPrime software. Interviewers followed identical administration instructions in all sites, set up each task for the adolescent, and allowed the adolescent to complete the task independently. The method of administration (i.e., in person or facilitated via Zoom) was consistent within each site, but not necessarily between sites. The fourth period of data collection occurred following the onset of the COVID-19 pandemic, during which participants completed tasks using a pre-packaged version of EPrime created during and for the pandemic (EPrime Go). Therefore, some sites were not able to collect data. For those sites that were able to participate (described in detail in the relevant sections of Results), adolescents completed EF tasks in one of three ways, largely dependent on COVID-19 restrictions in place at that site at the time of administration: (1) At their home or university location in a quiet outdoor setting (to mitigate COVID-19 risk) but in all other ways identical to the first three time periods; (2) over Zoom, where the adolescents received the instructions from the interviewer, and then took control of the interviewer's computer through a shared screen application; (3) self-administered on their own personal computer via a "packaged" version of the test battery created by the software company for use during the pandemic. The data were then uploaded via a secure server from the adolescent to the university that served as the data hub for the project.

2.3 | Measures

2.3.1 | EF

2.3.1.1 | ToL. The ToL (Berg and Byrd 2002) was used to measure how well adolescents could organize sequential behavior to reach a goal and if they could inhibit acting before a plan was fully formed. In this task, adolescents viewed a series of three balls on a peg in a start position that must be moved to a pre-specified goal configuration on three other pegs, one of which can support one ball, one of which can support two balls, and one of which can support three balls. The adolescent was instructed to replicate the goal configuration using the smallest number of moves. The number of perfectly solved trials for 3-, 4-, and 5-move problems was used as the ToL score for analyses (possible range of 0–4).

2.3.1.2 | Stroop. Stroop (1935) is a widely used measure of executive functions, requiring top-down internal control over behavior, attentional control, and inhibition or the ability to override a prepotent response. In this task, an individual was presented with a list of words printed in different colors and must identify the color in which a word is printed, while ignoring the meaning of the word (i.e., say "blue" when the word "red" is printed in blue ink). Adolescents received trials with color words in the same color ink (congruent) and trials with color words in a different color ink (incongruent). Performance was the proportion of accurate color naming for a set of 20 words during a block of mixed congruent and incongruent trials (possible range 0–1).

2.3.1.3 | WM. We used an item-recognition task to measure WM (Thompson-Schill et al. 2002) in which adolescents saw a target set of four letters on the computer screen. Then, a screen displaying a "+," followed by a screen displaying a probe letter was shown, and adolescents were asked whether the probe was a member of the current target set (to which they answered either "yes" or "no" using the keyboard). Half of the trials contained probe items that were members of the current target set (positive trials), and half of the trials contained probe items that were not members of the current target set (negative trials). For both positive and negative trials, half of the trials contained probe items that were members of the previous target set (hard/recent trials), and half of the trials contained probe items that were not members of either of the previous two target sets (easy/

non-recent trials). For all trials, two of the four letters in the target set were repeated from the previous trial (so that repetition of items in the target set was not confounded with trial type). After eight practice trials, adolescents were presented with a total of 64 trials (32 easy and 32 hard) in pseudo-random order, split into blocks of eight trials each. WM was measured as the number of correct responses on negative-recent trials (i.e., adolescents correctly indicated that the current probe was not among the previous target set; possible range 0–8).

2.4 | Data Analysis Plan

Descriptive statistics and bivariate correlations between main study variables were estimated, followed by latent growth curve modeling (LGM) to estimate the average intercepts and slopes and individual differences (i.e., variance) in the intercept and slope for each EF task. First, linear growth curves were estimated using the most complete dataset (all collected prior to the beginning of the COVID pandemic) involving the first three waves of EF data collection (at 10, 14, and 17 years of age on average). Analyses were conducted using M*plus* v.8.1 (Muthén and Muthén 2018). The fit of each model was assessed using RMSEA, CFI, SRMR, and model Chi-square (χ^2_M). The thresholds for good model fit as recommended by Kline (2023) were: RMSEA $\leq$ 0.05, CFI $\geq$ 0.90, SRMR $<$ 0.10, and a non-significant χ^2_M ($p > 0.05$). We then utilized meta-analytic methods for the linear model to test whether the average slope effect was homogenous across sites. These analyses were conducted in R using the "meta" package (Balduzzi et al. 2019; R Core Team 2021). For the random effect meta-analysis, we used the slope and intercept estimates from each country-specific model to estimate the pooled effect size for the full sample and the variability in this estimate across samples. To quantify the heterogeneity in effect sizes between the sites, we report τ^2 (i.e., the variance of the true effect sizes), I^2 (percentage of variability in the effect sizes that is not due to sampling error), and the prediction interval (the range of expected effect sizes) as recommended by Harrer et al. (2021). We use the following rule of thumb for interpreting the value of the I^2 statistic: 25% = low heterogeneity, 50% = moderate heterogeneity, 75% = substantial heterogeneity (Higgins and Thompson 2002). The τ^2 statistic is on the same scale as the effect size used, and therefore, the magnitude can be interpreted in the same way as a standard deviation.

Then, for the subset of study sites that were also able to collect a fourth wave of EF data, quadratic growth curves were estimated with available data from 10 to 21 years. We compared model fit between the linear and quadratic models for the full sample four-wave models using a Chi-square difference test ($\Delta\chi^2_M$). A significant p value would indicate that the quadratic model provided a better fit to the data than the linear model (Kline 2023). Table 1 presents a model fit comparison between linear and quadratic models for each task. The linear models were run first for the full sample and then for each individual study site. If a site had greater than 50% missingness on any measure across the three timepoints, a growth curve model was not estimated. Because of sample size constraints and missing data in the fourth EF data collection wave, most of the site-specific quadratic models did not converge, so we did not conduct meta-analysis or report quadratic model results separately by study site.

TABLE 1 | Model fit information for linear and quadratic four-wave growth curve models.

	Linear	Quadratic
Tower of London		
$\chi_M{}^2$ (df)	8.259 (8)	2.704 (4)
$\Delta\chi_M{}^2$ (Δdf)	—	5.55 (4)
Stroop		
$\chi_M{}^2$ (df)	105.804 (8)	34.988 (4)
$\Delta\chi_M{}^2$ (Δdf)	—	70.816 (4)***
Working memory		
$\chi_M{}^2$ (df)	271.080 (8)	73.728 (4)
$\Delta\chi_M{}^2$ (Δdf)	—	197.352 (4)***

$\chi_M{}^2$ = model Chi-square, df = degrees of freedom, $\Delta\chi_M{}^2$ = change in model Chi-square, Δdf = change in degrees of freedom.
***$p < 0.001$.

3 | Results

Independent samples t-tests revealed no differences in EF tasks between males and females (all $ps > 0.054$), except for a small difference ($d = 0.14$, $p = 0.026$, females > males) in Stroop performance at 14 years.

Table 2 presents correlations and descriptive statistics between all variables. Relatively normally distributed for Stroop at age 10 and ToL at ages 10, 14, and 17. For Stroop at ages 14 and 17, and WM at 10, 14, and 17, scores were somewhat negatively skewed or normal (−2.83 to −0.28); transformations were not used given the robustness of the general linear model to moderate degrees of skewness. Within each age, EF tasks were positively and significantly associated with each other. The means and distributions of scores on these tasks in our sample are similar to those reported in other samples (e.g., Injoque-Ricle et al. 2014; Khng and Lee 2014).

3.1 | Models for Three-Wave Data

Table 3 presents parameter estimates from the LGM results for the full sample and each site. Results are presented separately for each EF task. Figure 1 presents the growth curves for each task and site.

3.1.1 | ToL

Note that for one sample (Italy), there was more than 50% missingness on the ToL variable, so a site-specific model was not estimated for this location. The LGM for ToL for the full sample had a good fit to the data (RMSEA $<$ 0.001; SRMR = 0.007; χ^2_M = 0.54, $p = 0.462$, CFI = 1.00). Across the three timepoints, adolescents' ToL scores ranged from 0 to 4. On average, at age 10, adolescents' ToL score was 1.90 ($b = 1.90$, SE = 0.02, $p < 0.001$), and there was a significant positive slope for ToL from ages 10–17 ($b = 0.25$, SE = 0.02, $p < 0.001$). There were significant individual differences in adolescents' ToL performance at age 10 ($b = 0.13$, SE

TABLE 2 | Descriptive statistics and bivariate correlations.

	1	2	3	4	5	6	7	8	9	10	11	12
1. ToL 10	—											
2. ToL 14	0.27**	—										
3. ToL 17	0.18**	0.26**	—									
4. ToL 21	0.22**	0.28**	0.39**	—								
5. Stroop 10	0.18**	0.16**	−0.01	0.20**	—							
6. Stroop 14	0.21**	0.19**	0.10*	0.19**	0.33**	—						
7. Stroop 17	0.22**	0.17**	0.11**	−0.11	0.17**	0.26**	—					
8. Stroop 21	0.15*	0.09	0.07	0.18**	0.16**	0.06	0.03	—				
9. WM 10	0.29**	0.16**	0.07	0.22**	0.28**	0.30**	0.23**	0.09	—			
10. WM 14	0.21**	0.17**	0.14**	0.18**	0.24**	0.25**	0.24**	0.003	0.51**	—		
11. WM 17	0.14**	0.14**	0.16**	0.19**	0.07	0.17**	0.20**	0.03	0.32**	0.49**	—	
12. WM 21	0.15*	0.08	0.11	0.39**	0.16**	0.11	0.08	0.46**	0.09	0.003	0.28**	—
Mean	1.89	2.12	2.37	2.50	0.85	0.89	0.89	0.87	6.13	6.59	7.14	6.87
SD	0.68	0.67	0.69	0.77	0.14	0.15	0.16	0.19	1.41	1.29	0.97	1.61

Abbreviations: SD = standard deviation, ToL = Tower of London, WM = working memory task.

*$p < 0.05$.

**$p < 0.01$.

TABLE 3 | Linear growth curve estimates for executive function tasks.

Study site	Tower of London				Stroop				Working memory			
	Int. Mean (SE)	Slope Mean (SE)	Int. Variance (SE)	Slope Variance (SE)	Int. Mean (SE)	Slope Mean (SE)	Int. Variance (SE)	Slope Variance (SE)	Int. Mean (SE)	Slope Mean (SE)	Int. Variance (SE)	Slope Variance (SE)
Full sample	1.90 (0.02)***	0.25 (0.02)***	0.13 (0.04)***	0.03 (0.02)	0.86 (0.004)***	0.03 (0.003)***	0.01 (0.002)***	0.002 (0.001)*	6.11 (0.04)***	0.53 (0.02)***	1.23 (0.12)***	0.24 (0.04)***
Colombia	1.81 (0.08)***	0.45 (0.06)***	0.09 (0.04)*	N/A	0.84 (0.02)***	0.04 (0.01)**	0.01 (0.004)	0.001 (0.002)	4.97 (0.24)***	0.84 (0.13)***	1.06 (0.57)	0.02 (0.28)
Italy	N/A	N/A	N/A	N/A	0.90 (0.01)***	0.03 (0.01)***	0.01 (0.003)*	0.002 (0.001)*	6.73 (0.08)***	0.30 (0.04)***	0.49 (0.14)***	0.02 (0.06)
Jordan	1.68 (0.07)***	0.21 (0.05)***	0.11 (0.12)	0.05 (0.06)	0.80 (0.02)***	0.02 (0.01)	0.01 (0.01)	0.01 (0.004)	5.75 (0.13)***	0.54 (0.09)***	0.22 (0.30)	0.27 (0.15)
Kenya	1.24 (0.12)***	0.54 (0.07)***	0.62 (0.24)*	0.12 (0.12)	0.77 (0.03)***	−0.03 (0.02)	0.01 (0.01)	0.01 (0.01)	4.76 (0.27)***	1.03 (0.14)***	2.67 (1.10)*	0.75 (0.34)*
Philippines	2.04 (0.08)***	0.24 (0.06)***	0.10 (0.09)	0.05 (0.05)	0.86 (0.02)***	0.03 (0.01)	0.002 (0.003)	N/A	N/A	N/A	N/A	N/A
Thailand	2.05 (0.06)***	0.11 (0.04)*	0.13 (0.09)	0.01 (0.05)	0.90 (0.01)***	0.02 (0.01)**	0.01 (0.02)***	0.002 (0.001)*	N/A	N/A	N/A	N/A
United States	2.00 (0.06)***	0.23 (0.05)***	0.09 (0.12)	0.05 (0.07)	0.88 (0.01)***	0.04 (0.01)***	0.01 (0.004)	0.001 (0.003)	6.06 (0.13)***	0.46 (0.07)***	1.47 (0.37)***	0.06 (0.11)

Note: NA due to model estimation problems, the slope variance parameter in these models was constrained to 0. Tower of London—Italy and Working Memory—Philippines and Thailand models were not estimated due to missing data.

Abbreviations: Int. = intercept, N/A = not applicable, SE = standard error.

*$p < 0.05$.

***$p < 0.001$.

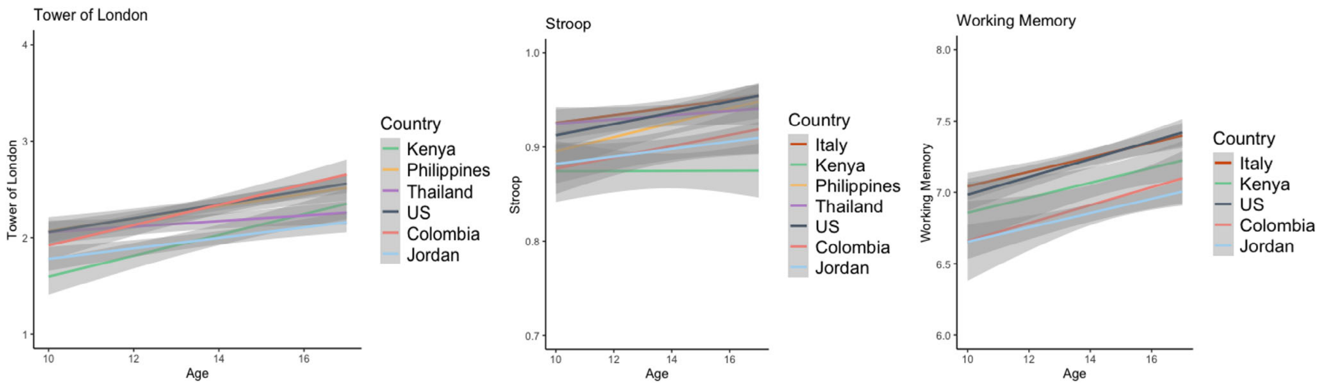

FIGURE 1 | Change in executive functioning variables from ages 10 to 17. *Note:* The gray area around each line indicates the standard error of the slope.

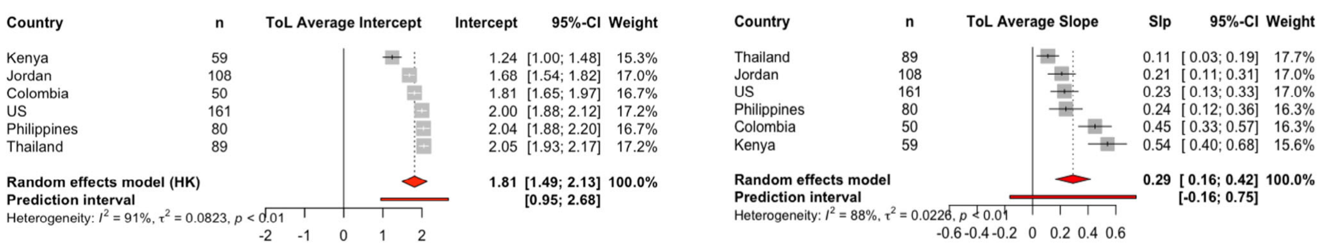

FIGURE 2 | Forest plot for Tower of London average slope and intercept. *Note:* There was significant between site heterogeneity in Tower of London performance at age 10 ($\tau^2 = 0.08$ (95% CI [0.03, 0.58]); $I^2 = 90.6\%$ (95% CI [82.3%, 95.0%]; prediction interval: ($g = 0.95$–2.68), and in Tower of London performance change from 10 to 17 ($\tau^2 = 0.02$ (95% CI [0.01, 0.16]); I^2 value of 87.9% (95% CI [76.1%, 93.9%]; prediction interval: $g = (-0.16$ to $0.75)$. CI = confidence interval, Slp = slope, ToL = Tower of London.

= 0.04, $p < 0.001$), but no significant individual differences in the slope of ToL performance from ages 10 to 17 ($b = 0.03$, SE = 0.02, $p = 0.134$). Table 3 presents estimates for each sample. For the separate samples, all samples had a significant, positive intercept for ToL at age 10, and three of the seven samples had a significant positive slope from ages 10 to 17.

The meta-analysis showed that for ToL, the pooled average intercept estimate was statistically significant ($b = 1.81$, $p < 0.001$). The between-site heterogeneity was estimated at $\tau^2 = 0.08$ (95% CI [0.03, 0.58]) with an I^2 value of 90.6% (95% CI [82.3%, 95.0%]), indicating substantial between-site heterogeneity. The prediction interval ranged from $g = 0.95$–2.68. The pooled average slope estimate was statistically significant ($b = 0.29$, $p < 0.001$). The between-site heterogeneity was estimated at $\tau^2 = 0.02$ (95% CI [0.01, 0.16]) with an I^2 value of 87.9% (95% CI [76.1%, 93.9%]), indicating substantial between-site heterogeneity. The prediction interval ranged from $g = -0.16$ to 0.75. Figure 2 displays the weighted effect sizes for the intercept and slope estimates for each country (study site).

3.1.2 | Stroop

The linear LGM for Stroop for the full sample had an adequate fit to the data (RMSEA = 0.110, SRMR = 0.047, $\chi^2_M = 16.36$, $p < 0.001$, CFI = 0.835). Across the three timepoints, adolescents' Stroop scores ranged from 0 to 1. On average, adolescents' Stroop performance at age 10 was 0.86 ($b = 0.86$, SE = 0.004, $p < 0.001$), and there was a significant linear increase in Stroop performance from ages 10 to 16 ($b = 0.03$, SE = 0.003, $p < 0.001$). There were significant individual differences in adolescents' Stroop performance at age 10 ($b = 0.01$, SE = 0.002, $p < 0.001$) and significant individual differences in the slope of Stroop performance from ages 10 to 17 ($b = 0.002$, SE = 0.001, $p = 0.024$). Table 3 presents estimates for each sample. All samples had a significant positive intercept, and four of the seven samples had a significant positive slope from ages 10 to 17 years. Of the three sites with nonsignificant mean slopes, two had a positive slope estimate.

The meta-analysis showed that for Stroop, the pooled average intercept estimate was statistically significant ($b = 0.85$, $p < 0.001$). The between-site heterogeneity was estimated at $\tau^2 = 0.002$ (95% CI [0.001, 0.01]) with an I^2 value of 85.8% (95% CI [72.8%, 92.6%]), indicating substantial between-site heterogeneity. The prediction interval ranged from $g = 0.73$–0.98. The pooled average slope estimate was statistically significant ($b = 0.03$, $p < 0.001$). The between-site heterogeneity was estimated at $\tau^2 < 0.001$ (95% CI [0.00, 0.003]) with an I^2 value of 52.5% (95% CI [0%, 79.8%]), indicating moderate between-site heterogeneity. The prediction interval ranged from $g = 0.002$–0.05. Figure 3 displays the weighted effect sizes for the intercept and slope for each country (study site).

3.1.3 | WM

Note that for two samples (Thailand and the Philippines), there was more than 50% missingness on the WM variable, so

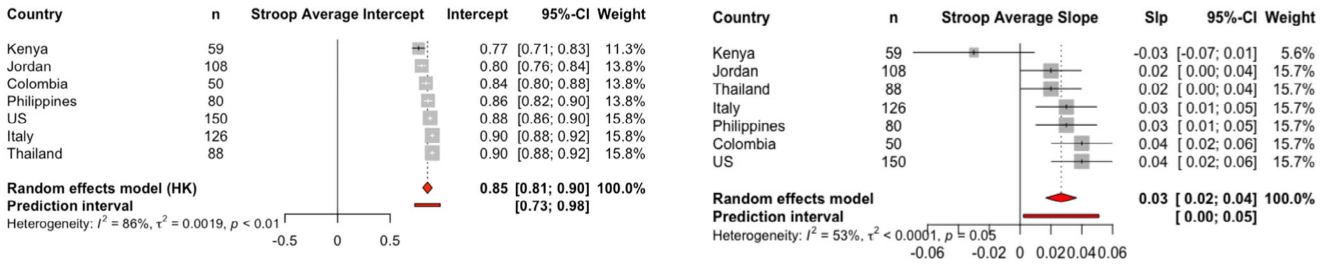

FIGURE 3 | Forest plot for Stroop average slope and intercept. *Note:* There was significant between site heterogeneity in Stroop performance at age 10 (τ^2 = 0.002 (95% CI [0.001, 0.01]); I^2 = 85.8% (95% CI [72.8%, 92.6%]); prediction interval: g = (0.73–0.98), and small to moderate between-site heterogeneity in Stroop performance change from 10 to 17 (τ^2 < 0.001 (95% CI [0.00, 0.003]); I^2 value of 52.5% (95% CI [0%, 79.8%]); prediction interval: g = (0.002–0.05). CI = confidence Interval, Slp = slope.

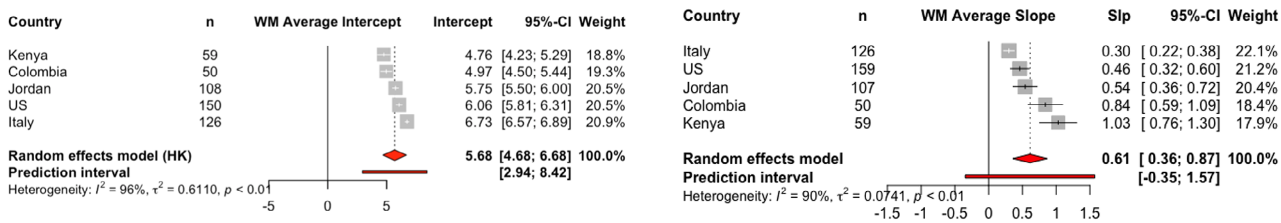

FIGURE 4 | Forest plot for Working Memory average slope and intercept. *Note:* There was significant between site heterogeneity in Working Memory performance at age 10 (τ^2 = 0.61 (95% CI [0.20, 5.33]); I^2 = 96.4% (95% CI [93.9%, 97.9%]); prediction interval: (g = 2.94–8.42), and in Working Memory performance change from 10 to 17 (τ^2 = 0.07 (95% CI [0.02, 0.71]); I^2 = 90.1% (95% CI [79.8%, 95.2%]); prediction interval: g = −0.35 to 1.57. CI = confidence interval, Slp = slope, WM = working memory.

site-specific models were not estimated for those two locations. The linear LGM for WM for the full sample had a good fit to the data (RMSEA = 0.093; SRMR = 0.027; χ^2_M = 12.18, p < 0.001, CFI = 0.975). Across the three timepoints, adolescents' WM scores ranged from 0 to 8. On average, at age 10, adolescents' WM score was 6.11 (b = 6.11, SE = 0.04, p < 0.001), and there was a significant positive slope for WM from ages 10 to 17 (b = 0.53, SE = 0.02, p < 0.001). There were significant individual differences in adolescents' WM performance at age 10 (b = 1.23, SE = 0.12, p < 0.001) and significant individual differences in the slope of WM performance from ages 10 to 17 years (b = 0.24, SE = 0.04, p < 0.001). Table 3 presents estimates for each sample. All five samples that we analyzed showed a significant positive intercept and a significant positive slope.

The meta-analysis showed that for WM, the pooled average intercept estimate was statistically significant (b = 5.68, p < 0.001). The between-site heterogeneity was estimated at τ^2 = 0.61 (95% CI [0.20, 5.33]) with an I^2 value of 96.4% (95% CI [93.9%, 97.9%]), indicating substantial between-site heterogeneity. The prediction interval ranged from g = 2.94–8.42. The pooled average slope estimate was statistically significant (b = 0.61, p < 0.001). The between-site heterogeneity was estimated at τ^2 = 0.07 (95% CI [0.02, 0.71]) with an I^2 value of 90.1% (95% CI [79.8%, 95.2%]), indicating substantial between-site heterogeneity. The prediction interval ranged from g = −0.35 to 1.57. Figure 4 displays the weighted effect sizes for the intercept and slope for each study site.

3.2 | Models for Four-Wave Data

Before proceeding with analysis of the four-wave dataset, we examined the fourth wave data (collected when adolescents were 21 years old on average, after the COVID pandemic had begun) to see whether individual differences in EF performance were associated with self-reports of a positive test for COVID-19 or severity of COVID-19 symptoms on an ordinal scale. None of the estimated associations was significant (p values from 0.10 to 0.44), so we did not control for COVID test status in subsequent analyses. Figure 5 presents visualizations for linear and quadratic change in the three tasks for the full sample.

3.2.1 | ToL

The sample for the quadratic growth curve for ToL from ages 10 to 21 years consisted of individuals from Jordan, Philippines, and the United States. The Chi-square difference test between the linear and quadratic models indicated that the quadratic model did not significantly improve model fit ($\Delta\chi^2_M$ = 8.26, Δdf = 4, p = 0.235). Thus, we present results from the linear four-wave model. On average, at age 10, adolescents' ToL score was estimated at 1.91 (b = 1.91, SE = 0.02, p < 0.001), and there was a significant positive slope for ToL from ages 10 to 21 years (b = 0.22, SE = 0.01, p < 0.001). There were significant individual differences in adolescents' ToL performance at age 10 (b = 0.13, SE = 0.03, p < 0.001) and in the slope of ToL performance from ages 10 to 21 years (b = 0.02, SE = 0.01, p = 0.004). Table 4 presents estimates from the linear model.

3.2.2 | Stroop

The sample for the quadratic growth curve for Stroop from 10 to 21 years consists of individuals from Colombia, Italy, Jordan, Philippines, and the United States. The Chi-square difference

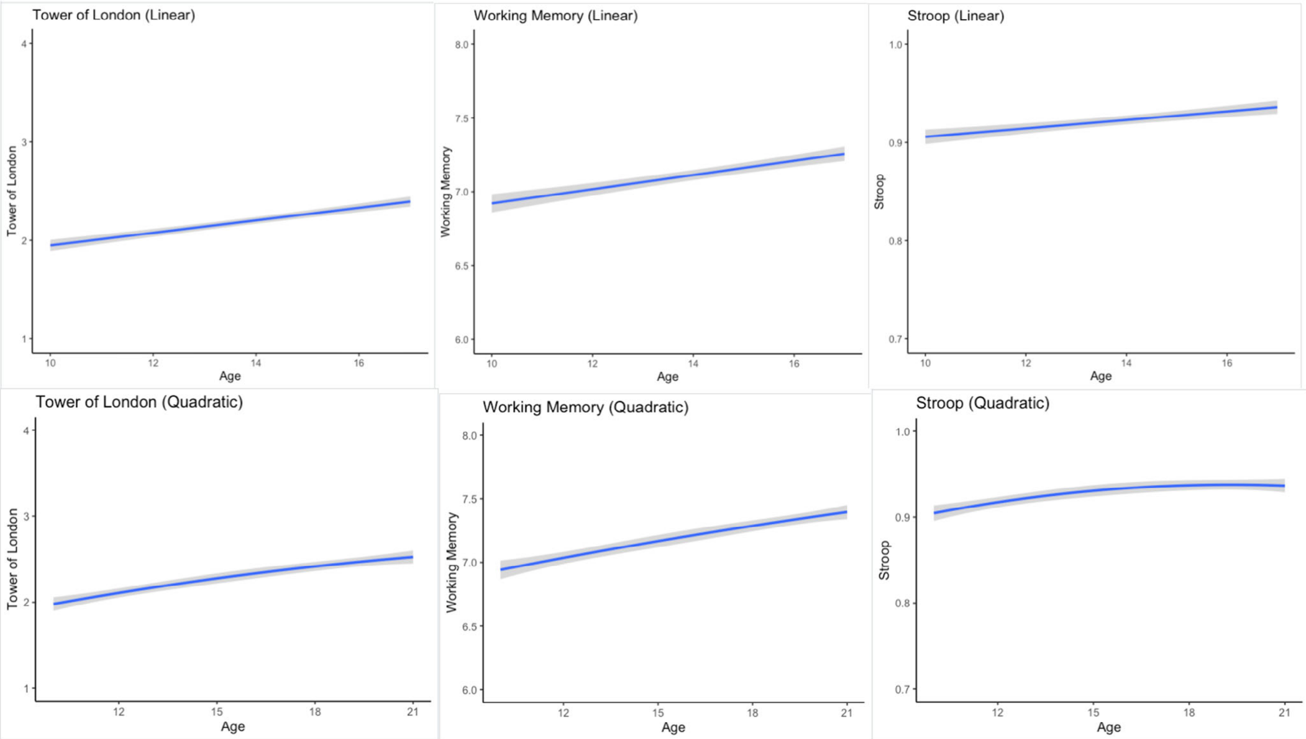

FIGURE 5 | Linear (top row) and quadratic (bottom row) change for each task in the full sample. *Note:* For the Tower of London, the quadratic model did not significantly improve model fit. For Working Memory and Stroop, the quadratic model did significantly improve model fit.

test between the linear and quadratic models indicated that the quadratic model significantly improved model fit ($\Delta\chi^2_M$ = 105.80, $\Delta df = 4$, $p < 0.001$). Thus, we present results from the quadratic four-wave model. On average, at age 10, adolescents' Stroop score was estimated at 0.85 ($b = 0.85$, SE = 0.01, $p <$ 0.001), and there was a significant positive linear slope for Stroop from ages 10 to 21 years ($b = 0.06$, SE = 0.01, $p < 0.001$), and a significant negative quadratic slope ($b = -0.02$, SE = 0.003, p < 0.001). Variance in adolescents' Stroop performance at age 10 was marginally significant ($b = 0.003$, SE = 0.002, $p = 0.061$), but the linear slope of Stroop performance from ages 10 to 21 years was not significant ($b = 0.01$, SE = 0.01, $p = 0.169$). However, significant individual differences in the quadratic term emerged ($b = 0.002$, SE = 0.001, $p = 0.004$). Table 4 presents all estimates from the quadratic model. However, we interpret these results with caution because the analyses included warning messages (i.e., non-positive definite latent variable covariance matrix).

3.2.3 | WM

The sample for the quadratic growth curve for WM from ages 10 to 21 years consists of individuals from Colombia, Italy, Jordan, and the United States. The Chi-square difference test between the linear and quadratic models indicated that the quadratic model significantly improved model fit ($\Delta\chi^2_M$ = 197.35, Δdf = 4, $p < 0.001$). Thus, we present results from the quadratic four-wave model. On average, at age 10, adolescents' WM score was estimated at 6.06 ($b = 6.06$, SE = 0.06, $p < 0.001$), and there was a significant positive linear slope for WM from ages 10 to 21 (b = 0.91, SE = 0.07, $p < 0.001$), and a significant negative quadratic slope ($b = -0.20$, SE = 0.02, $p < 0.001$). There were significant individual differences in adolescents' WM performance at age 10 ($b = 1.49$, SE = 0.13, $p < 0.001$), the linear slope of WM performance from ages 10 to 21 years ($b = 0.75$, SE = 0.20, $p <$ 0.001), and the quadratic term ($b = 0.12$, SE = 0.02, $p < 0.001$). Table 4 presents all estimates from the quadratic model.

4 | Discussion

To our knowledge, the current study is the first to report longitudinal growth in EF across adolescence into young adulthood in an international sample. Based on our models, we found that EF task performance (i.e., ToL, Stroop, and WM) improved linearly from age 10 to 14 to 17 years (average ages at each of the three assessment waves). This finding extends prior longitudinal work spanning early and middle childhood, showing average linear improvements in EF (Blankenship et al. 2019; Castellanos-Ryan et al. 2023). However, based on the current study's fourth wave of data for a subset of study sites (spanning 10–21 years), EF performance growth on two of the three tasks decelerated in the transition across late adolescence and young adulthood. Although this finding from the current study should be interpreted with caution because the fourth wave of data was collected in the era of the COVID pandemic, the result is consistent with recent rigorous cross-sectional and longitudinal research that points to deceleration in growth of EF performance at the same developmental period when moving from the teens into the 20s (e.g., Ferguson et al. 2021; Tervo-Clemmens et al. 2023).

Turning from averages to individual variation, the observed variances in EF task performance were largely accounted for by the observed variation at the first assessment (10 years of age on average). For all three tasks, results from the growth curve models showed significant and substantial individual difference variance in the intercept (i.e., task performance at age 10), and modest (but still usually significant) variance in the slope. These findings accord with prior research in childhood showing that EF shows strong rank order stability that is moderately to substantially consistent over time (e.g., Deater-Deckard 2014; Gooch et al. 2016; Willoughby et al. 2012). Thus, individuals who are lower or higher in performance relative to their peers in the transition to early adolescence generally maintain that rank order position later across adolescent development, although there still remain some individual differences in how slowly or rapidly their EF performance improves over time.

Were the intercept and linear growth slope estimates consistent across the international study sites? To answer this question, we used meta-analysis, treating each site as a distinct study (Folker et al. 2023). For all three EF tasks, the meta-analyzed pooled effect sizes for average slope were statistically significant, indicating that across the international sites, task performance significantly increased from 10 to 14 to 17 years. This confirmed the finding from the total-sample latent growth model results reported above. One advantage of the meta-analysis approach is that it also quantifies the magnitude of between-study (in our case, study site) heterogeneity in model estimates. For all three EF tasks, there was substantial between-site heterogeneity in the slope and intercept estimates. Thus, these international study sites differed in their starting points and rates of change in EF task performance.

In light of the substantial between-site variability in intercept and slope estimates, were there any discernible patterns with regard to country differences? In short, "no." For example, the adolescents in Kenya showed the linear slope closest to 0 (i.e., no growth) compared to other sites for the Stroop task, but the strongest linear slope compared to the other sites for ToL and WM. Additionally, we examined whether patterns emerged when grouping sites according to national gross domestic product (GDP) as done in Duell et al. (2016), and saw no consistent patterns emerge for intercepts and slopes across the three EF tasks. The study samples are not nationally representative, so the results do not generalize at the national or international level. What we can say with confidence is that, within the sampling constraints of this large and diverse international sample, there was substantial heterogeneity in EF development within and between study sites. Future research will do well to investigate the potential correlates and causes of these multiple levels of variation.

EF varies widely and consistently between individuals and is a crucial foundation for academic and social-emotional functioning and health in adolescence and later developmental periods (Cartwright 2012; Yang et al. 2022). Consequently, EF has been deemed an important target for interventions seeking to improve health outcomes and mitigate deleterious outcomes. For example, interventions targeting EF skills (i.e., using games to target WM and inhibition) in childhood are effective in improving academic outcomes (Mattera et al. 2021). Moreover, Klingberg et al. (2005) used an intervention to improve EF in children ages 7–12 with ADHD and found that enhanced EF skills were associated with less impulsive behavior. Meta-analyses by Allom et al. (2016) and

TABLE 4 | Growth curve estimates for quadratic models.

	Int. mean (SE)	Linear slope mean (SE)	Quadratic slope mean (SE)	Int. variance (SE)	Linear slope variance (SE)	Quadratic slope variance (SE)
Tower of London	1.91 (0.02)***	0.22 (0.01)***	N/A	0.13 (0.03)***	0.02 (0.01)**	N/A
Stroop	0.85 (0.01)***	0.06 (0.01)***	−0.02 (0.003)***	0.003 (0.002)	0.01 (0.01)	0.002 (0.001)***
Working memory	6.06 (0.06)***	0.91 (0.07)***	−0.20 (0.02)***	1.49 (0.13)***	0.75 (0.20)***	0.12 (0.02)***

Abbreviations: Int. = intercept, N/A = not applicable, SE = standard error.
** $p < 0.01$.
*** $p < 0.001$.

Protogerou et al. (2020) showed that inhibitory/self-control and impulsivity training interventions are sometimes associated with reduced risk-taking and harmful health-risk behaviors. However, as these and other meta-analyses of EF skills have noted, effect sizes generally are small, and often attenuate over time and do not generalize to other and broader domains of functioning beyond the targeted domain of training (e.g., Kassai et al. 2019; Scionti et al. 2020).

Two new pieces of information from the current study add to ongoing considerations of targeting EF skills for intervention. First, regarding developmental timing, the current results showed substantial variation between individuals in the first assessment time point (age 10 years on average) and modest variation in growth trajectories beyond 10 years of age. Second, regarding within- and between-country differences, the current results showed moderate to substantial heterogeneity in average EF performance at age 10 years and average growth trajectories across adolescence. This result would suggest that, across a variety of geospatial and cultural contexts, adolescence still may be a developmental period of plasticity for EF intervention. There are environmental (i.e., household chaos, socioeconomic status) and biological (i.e., puberty) factors that have been shown to be associated with individual differences in EF during childhood and early adolescence (Albert et al. 2020; Gerván et al. 2024; Vernon-Feagans et al. 2016). Although it was outside of the scope of this study to examine potential predictors of differences in EF intercept and slope between individuals and study sites, future work may consider investigating the role of these factors and how they may explain variation in EF.

There are a number of limitations to consider. First, our capacity to estimate nonlinear change in EF task performance was constrained by the smaller sample size overall and the loss of several study sites in the fourth wave assessment (due largely to the impact of the COVID-19 pandemic). Therefore, we were unable to test whether linear or non-linear change fit the data better within each site. Second, longitudinal assessments of EF every 3–4 years provided minimally sufficient data for estimating growth trajectories. However, having more frequent assessments would be advantageous in at least two ways: (1) they would permit more precise estimation of individual differences in linear and nonlinear change and (2) they would allow for investigation of potential latent growth classes that would more effectively distinguish subgroups of individuals who show slower versus rapid improvements in EF. Third, the study sites' samples were intended to be representative of the local community, but are not representative of each nation. The results may not generalize to samples that are nationally representative. Fourth, given that the aims of the present study were to estimate average change and to see whether that change was homogeneous across diverse international samples, we did not examine correlates that might explain the between-site heterogeneity of effects. Future work will extend the current analyses to identify variables that differ on average between study sites that could help explain some between-site heterogeneity in EF task performance and growth across adolescence. Those analyses can also take into account the statistical prediction of within-person change in EF performance.

In conclusion, the present study contributes to the current EF literature by showing that EF task performance improves linearly and non-linearly, such that the rate of improvement slows for Stroop and WM from ages 10 to 21 years of age in diverse cultural and national contexts. There was noteworthy between-site variation in the estimated starting points and growth trajectories in EF across adolescence, but the study site differences were not uniform across EF tasks or growth parameters. It is important for future studies to consider individual- and group-level factors contributing to variation in EF task performance in the transition to and across adolescence. Prior cross-sectional work has shown that SES, parenting behaviors, household chaos, and other family environment factors are associated with EF (Andrews et al. 2021; Bernier et al. 2015; Kao et al. 2018)—although we do not know of any longitudinal work examining whether such factors are associated with changes in EF across adolescence. Addressing this gap in knowledge will be important to advancing our understanding of the growth and development of EF throughout adolescence and after, which may lead to more effective promotion of optimizing environments and improving outcomes.

Acknowledgments

We are grateful to the families who shared their time and participation with us, the Parenting Across Cultures group, and our community partners.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Code for the analyses is available upon request.

Preregistration

This study was not preregistered.

References

Albert, W. D., J. L. Hanson, A. T. Skinner, et al. 2020. "Individual Differences in Executive Function Partially Explain the Socioeconomic Gradient in Middle-School Academic Achievement." *Developmental Science* 23, no. 5: e12937.

Allom, V., B. Mullan, and M. Hagger. 2016. "Does Inhibitory Control Training Improve Health Behaviour? A Meta-Analysis." *Health Psychology Review* 10, no. 2: 168–186. https://doi.org/10.1080/17437199.2015.1051078.

Anderson, V., P. J. Anderson, R. Jacobs, and M. S. Smith. 2010. "Development and Assessment of Executive Function: From Preschool to Adolescence." In *Executive Functions and the Frontal Lobes*, 157–188. Psychology Press.

Andrews, K., L. Atkinson, M. Harris, and A. Gonzalez. 2021. "Examining the Effects of Household Chaos on Child Executive Functions: A Meta-Analysis." *Psychological Bulletin* 147, no. 1: 16–32. https://doi.org/10.1037/bul0000311.

Balduzzi, S., G. Rücker, and G. Schwarzer. 2019. "How to Perform a Meta-Analysis With R: A Practical Tutorial." *Evidence Based Mental Health* 22, no. 4: 153–160. https://doi.org/10.1136/ebmental-2019-300117.

Barrett, H. C. 2022. "Psychology Within and Without the State." *Annual Review of Psychology* 73: 461–487. https://doi.org/10.1146/annurev-psych-020821-110248.

Berg, W. K., and D. L. Byrd. 2002. "The Tower of London Spatial Problem-Solving Task: Enhancing Clinical and Research Implementation." *Journal of Clinical and Experimental Neuropsychology* 24, no. 5: 586–604. https://doi.org/10.1076/jcen.24.5.586.1006.

Bernier, A., M. H. Beauchamp, S. M. Carlson, and G. Lalonde. 2015. "A Secure Base From Which to Regulate: Attachment Security in Toddlerhood as a Predictor of Executive Functioning at School Entry." *Developmental Psychology* 51, no. 9: 1177–1189. https://doi.org/10.1037/dev0000032.

Blankenship, T. L., M. A. Slough, S. D. Calkins, K. Deater-Deckard, J. Kim-Spoon, and M. A. Bell. 2019. "Attention and Executive Functioning in Infancy: Links to Childhood Executive Function and Reading Achievement." *Developmental Science* 22, no. 6: 12824. https://doi.org/10.1111/desc.12824.

Cartwright, K. B. 2012. "Insights From Cognitive Neuroscience: The Importance of Executive Function for Early Reading Development and Education." *Early Education & Development* 23, no. 1: 24–36. https://doi.org/10.1080/10409289.2011.615025.

Castellanos-Ryan, N., S. Parent, S. Chaput-Langlois, et al. 2023. "Modelling Executive Function Across Early Childhood: Longitudinal Invariance, Development From 3.5 to 7 Years and Later Academic Performance." *Cognitive Development* 68: 101365. https://doi.org/10.1016/j.cogdev.2023.101365.

Core Team R. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Cortés Pascual, A., N. Moyano Muñoz, and A. Quílez Robres. 2019. "The Relationship Between Executive Functions and Academic Performance in Primary Education: Review and Meta-Analysis." *Frontiers in Psychology* 10: 1582. https://doi.org/10.3389/fpsyg.2019.01582.

Deater-Deckard, K. 2014. "Family Matters: Intergenerational and Interpersonal Processes of Executive Function and Attentive Behavior." *Current Directions in Psychological Science* 23, no. 3: 230–236. https://doi.org/10.1177/0963721414531597.

Diamond, A. 2016. "*Why Improving and Assessing Executive Functions Early in Life Is Critical*." Edited by J. A. Griffin P. McCardle, and L. S. Freund, 11–43. American Psychological Association. https://doi.org/10.1037/14797-002.

Duell, N., L. Steinberg, J. Chein, et al. 2016. "Interaction of Reward Seeking and Self-Regulation in the Prediction of Risk Taking: A Cross-National Test of the Dual Systems Model." *Developmental Psychology* 52, no. 10: 1593–1605. https://doi.org/10.1037/dev0000152.

Erkut, S. 2010. "Developing Multiple Language Versions of Instruments for Intercultural Research." *Child Development Perspectives* 4, no. 1: 19–24. https://doi.org/10.1111/j.1750-8606.2009.00111.x.

Ferguson, H. J., V. E. Brunsdon, and E. E. Bradford. 2021. "The Developmental Trajectories of Executive Function From Adolescence to Old Age." *Scientific Reports* 11, no. 1: 1382. https://doi.org/10.1038/s41598-020-80866-1.

Folker, A. E., K. Deater-Deckard, J. E. Lansford, et al. 2023. "Intraindividual Variability in Parental Acceptance-Rejection Predicts Externalizing and Internalizing Symptoms Across Childhood/Adolescence in Nine Countries." *Journal of Family Psychology* 38, no. 2: 333–344. https://doi.org/10.1037/fam0001133.

Friedman, N. P., and A. Miyake. 2017. "Unity and Diversity of Executive Functions: Individual Differences as a Window on Cognitive Structure." *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior* 86: 186–204. https://doi.org/10.1016/j.cortex.2016.04.023.

Gerván, P., G. Oláh, K. Utczás, et al. 2024. "The Influence of Relative Pubertal Maturity on Executive Function Development in Adolescent Girls." *Scientific Reports* 14: 28140.

Gooch, D., P. Thompson, H. M. Nash, M. J. Snowling, and C. Hulme. 2016. "The Development of Executive Function and Language Skills in the Early School Years." *Journal of Child Psychology and Psychiatry* 57, no. 2: 180–187. https://doi.org/10.1111/jcpp.12458.

Harrer, M., P. Cuijpers, T. Furukawa, and D. Ebert. 2021. *Doing Meta-Analysis With R: A Hands-on Guide*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347.

Higgins, J. P., and S. G. Thompson. 2002. "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in Medicine* 21, no. 11: 1539–1558. https://doi.org/10.1002/sim.1186.

Hofmann, W., B. J. Schmeichel, and A. D. Baddeley. 2012. "Executive Functions and Self-Regulation." *Trends in Cognitive Sciences* 16, no. 3: 174–180. https://doi.org/10.1016/j.tics.2012.01.006.

Icenogle, G., L. Steinberg, N. Duell, et al. 2019. "Adolescents' Cognitive Capacity Reaches Adult Levels Prior to Their Psychosocial Maturity: Evidence for a "Maturity Gap" in a Multinational, Cross-Sectional Sample." *Law and Human Behavior* 43, no. 1: 69–85. https://doi.org/10.1037/lhb0000315.

Injoque-Ricle, I., J. P. Barreyro, A. Calero, and D. I. Burin. 2014. "Tower of London: Planning Development in Children From 6 to 13 Years of Age." *Spanish Journal of Psychology* 17: E77. https://doi.org/10.1017/sjp.2014.83.

Jacobson, L. A., A. P. Williford, and R. C. Pianta. 2011. "The Role of Executive Function in Children's Competent Adjustment to Middle School." *Child Neuropsychology* 17, no. 3: 255–280. https://doi.org/10.1080/09297049.2010.535654.

Kassai, R., J. Futo, Z. Demetrovics, and Z. K. Takacs. 2019. "A Meta-Analysis of the Experimental Evidence on the Near- and Far-Transfer Effects Among Children's Executive Function Skills." *Psychological Bulletin* 145, no. 2: 165–188. https://doi.org/10.1037/bul0000180.

Kao, K., S. Nayak, S. N. Doan, and A. R. Tarullo. 2018. "Relations Between Parent EF and Child EF: The Role of Socioeconomic Status and Parenting on Executive Functioning in Early Childhood." *Translational Issues in Psychological Science* 4, no. 2: 122–137. https://doi.org/10.1037/tps0000154.

Khng, K. H., and K. Lee. 2014. "The Relationship Between Stroop and Stop-Signal Measures of Inhibition in Adolescents: Influences From Variations in Context and Measure Estimation." *PLoS ONE* 9, no. 7: e101356. https://doi.org/10.1371/journal.pone.0101356.

Klimkeit, E. I., J. B. Mattingley, D. M. Sheppard, M. Farrow, and J. L. Bradshaw. 2004. "Examining the Development of Attention and Executive Functions in Children With a Novel Paradigm." *Child Neuropsychology* 10, no. 3: 201–211. https://doi.org/10.1080/09297040409609811.

Kline, R. B. 2023. *Principles and Practice of Structural Equation Modeling*. Guilford publications.

Klingberg, T., E. Fernell, P. J. Olesen, et al. 2005. "Computerized Training of Working Memory in Children With ADHD-A Randomized, Controlled Trial." *Journal of the American Academy of Child & Adolescent Psychiatry* 44, no. 2: 177–186.

Lansford, J. E., C. Sharma, P. S. Malone, et al. 2014. "Corporal Punishment, Maternal Warmth, and Child Adjustment: A Longitudinal Study in Eight Countries." *Journal of Clinical Child & Adolescent Psychology* 43, no. 4: 670–685. https://doi.org/10.1080/15374416.2014.893518.

Lee, K., R. Bull, and R. M. Ho. 2013. "Developmental Changes in Executive Functioning." *Child Development* 84, no. 6: 1933–1953.

Magar, E. C., L. H. Phillips, and J. A. Hosie. 2010. "Brief Report: Cognitive Regulation Across the Adolescent Years." *Journal of Adolescence* 33, no. 5: 779–781. https://doi.org/10.1016/j.adolescence.2009.10.002.

Mattera, S., N. M. Rojas, P. A. Morris, and K. Bierman. 2021. "Promoting EF With Preschool Interventions: Lessons Learned From 15 Years of Conducting Large-Scale Studies." *Frontiers in Psychology* 12: 640702.

Maxwell, B. 1996. "Translation and Cultural Adaptation of the Survey Instruments." *Third International Mathematics and Science Study (TIMSS) Technical Report* 1: 159–169.

Muthén, L. K., and B. O. Muthén. 2018. *Mplus User's Guide*. 8th ed. (Original work published 1998). Muthén & Muthén.

Nielsen, M., D. Haun, J. Kärtner, and C. H. Legare. 2017. "The Persistent Sampling Bias in Developmental Psychology: A Call to Action." *Journal of Experimental Child Psychology* 162: 31–38. https://doi.org/10.1016/j.jecp.2017.04.017.

Nyongesa, M. K., D. Ssewanyana, A. M. Mutua, et al. 2019. "Assessing Executive Function in Adolescence: A Scoping Review of Existing Measures and Their Psychometric Robustness." *Frontiers in Psychology* 10: 311. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00311.

Peña, E. D. 2007. "Lost in Translation: Methodological Considerations in Cross-Cultural Research." *Child Development* 78, no. 4: 1255–1264. https://doi.org/10.1111/j.1467-8624.2007.01064.x.

Protogerou, C., R. K. McHugh, and B. T. Johnson. 2020. "How Best to Reduce Unhealthy Risk-Taking Behaviours? A Meta-Review of Evidence Syntheses of Interventions Using Self-Regulation Principles." *Health Psychology Review* 14, no. 1: 86–115.

Raffaelli, M., L. J. Crockett, and Y. L. Shen. 2005. "Developmental Stability and Change in Self-Regulation From Childhood to Adolescence." *Journal of Genetic Psychology* 166, no. 1: 54–76. https://doi.org/10.3200/GNTP.166.1.54-76.

Scionti, N., M. Cavallero, C. Zogmaister, and G. M. Marzocchi. 2020. "Is Cognitive Training Effective for Improving Executive Functions in Preschoolers? A Systematic Review and Meta-Analysis." *Frontiers in Psychology* 10: 2812.

Shallice, T. 1982. "Specific Impairments of Planning." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 298, no. 1089: 199–209.

Steinberg, L. 2005. "Cognitive and Affective Development in Adolescence." *Trends in Cognitive Sciences* 9, no. 2: 69–74. https://doi.org/10.1016/j.tics.2004.12.005.

Steinberg, L. 2017. "A Social Neuroscience Perspective on Adolescent Risk-Taking." *In Biosocial Theories of Crime*, 435–463. Routledge.

Steinberg, L., G. Icenogle, E. P. Shulman, et al. 2018. "Around the World, Adolescence Is a Time of Heightened Sensation Seeking and Immature Self-Regulation." *Developmental Science* 21, no. 2: e12532. https://doi.org/10.1111/desc.12532.

Stroop, J. R. 1935. "Studies of Interference in Serial Verbal Reactions." *Journal of Experimental Psychology* 18, no. 6: 643–662.

Tervo-Clemmens, B., F. Calabro, A. Parr, J. Fedor, W. Foran, and B. Luna. 2023. "A Canonical Trajectory of Executive Function Maturation From Adolescence to Adulthood." *Nature Communications* 14: 6922. https://doi.org/10.1038/s41467-023-42540-8.

Theodoraki, T. E., S. P. McGeown, S. M. Rhodes, and S. E. MacPherson. 2020. "Developmental Changes in Executive Functions During Adolescence: A Study of Inhibition, Shifting, and Working Memory." *British Journal of Developmental Psychology* 38, no. 1: 74–89. https://doi.org/10.1111/bjdp.12307.

Thompson-Schill, S. L., J. Jonides, C. Marshuetz, et al. 2002. "Effects of Frontal Lobe Damage on Interference Effects in Working Memory." *Cognitive, Affective, & Behavioral Neuroscience* 2, no. 2: 109–120. https://doi.org/10.3758/CABN.2.2.109.

Vernon-Feagans, L., M. Willoughby, and P. Garrett-Peters. 2016. "Predictors of Behavioral Regulation in Kindergarten: Household Chaos, Parenting, and Early Executive Functions." *Developmental Psychology* 52, no. 3: 430.

Wechsler, D. 1945. "A Standardized Memory Scale for Clinical Use." *Journal of Psychology* 19, no. 1: 87–95.

Willoughby, M. T., R. J. Wirth, and C. B. Blair, & Family Life Project Investigators. 2012. "Executive Function in Early Childhood: Longitudinal Measurement Invariance and Developmental Change." *Psychological Assessment* 24, no. 2: 418–431. https://doi.org/10.1037/a0025779.

Yang, Y., G. S. Shields, Y. Zhang, H. Wu, H. Chen, and A. L. Romer. 2022. "Child Executive Function and Future Externalizing and Internalizing Problems: A Meta-Analysis of Prospective Longitudinal Studies." *Clinical Psychology Review* 97: 102194. https://doi.org/10.1016/j.cpr.2022.102194.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Table S1: Participant sex assigned at birth and age by country.